An Interop Publication

CONNEXIONS

The Interoperability Report

November 1994 Volume 8, No. 11

ConneXions — The Interoperability Report tracks current and emerging standards and technologies within the computer and communications industry.

In this issue:

ConneXions is published monthly by Interop Company, a division of Softbank Exposition and Conference Company, 303 Vintage Park Drive, Foster City, California, 94404–1138, USA.
Phone: +1 (415) 578-6900
Fax: +1 (415) 525-0194
E-mail: `connexions@interop.com`

Subscription hotline: 1-800-575-5717
or +1-502-493-3217

ISSN 0894–5926

From the Editor

A few events during the final days of October and early days of November warrant mention in this editorial. *NetWorld+Interop 94* took place in Paris and attracted more than 35,000 visitors. On the final day of the show it was announced that Ziff-Davis Publishing Company had been sold to the New York investment firm of Forstmann Little & Company. This sale did *not* however include ZD Expos, of which *ConneXions* is a part. The announcement that *our* new owners will be Softbank Corporation of Tokyo came a few days later. (See page 31). The new name for ZD Expos is "Softbank Exposition and Conference Company." Our editorial offices will remain in Foster City, and I do not expect any changes in editorial policy. I am also happy to say that our new subscription agency, The Cobb Group, will continue to provide customer service, even though we are no longer part of the same family of companies.

The *Internet Domain Survey* by Mark Lottor of Network Wizards of Menlo Park was also released in late October, and the numbers are very impressive. The total number of Internet hosts now stands at 3.8 million, with a 3rd quarter 1994 overall growth of 21 percent. The projected point at which the number of Internet hosts will hit 100 million is now 1st quarter 1999. The `.com` (commercial) domain is now the largest, and grew by a striking 36 percent during the quarter to 1,054,422 hosts. Also, the `.net` (network) domain, which is heavily used by many public data Internet service providers for their customers, grew by 66 percent during the same period.

As reported in our August issue, the current NSFNET backbone will eventually be replaced by a system of interconnected network providers. This month, Jessica Yu, Enke Chen, and Laurent Joncheray of Merit Network Inc., describe a proposed routing solution for the initial system of *Network Access Points* (NAPs).

Our second article, by Walter Willinger, Daniel V. Wilson, Will E. Leland, and Murad S. Taqqu takes a hard look at traditional methods for modeling traffic in computer networks. The authors introduce the concept of *self-similar* or *fractal* models for high-speed network traffic, and present preliminary evidence for how these new models impact practically all aspects of network engineering.

With the advent of *Multi-Media Internet Mail Extensions* (MIME) it is possible to send e-mail messages which contain encoded voice. Our final article by Greg Vaudreuil of Octel Communications is a look at work underway to develop a MIME profile for voice messaging.

A Routing Design for the Initial ATM NAP Architecture

by Jessica Yu, Enke Chen, and Laurent Joncheray, Merit Network, Inc.

Abstract

The initial ATM NAP architecture planned by several NAP providers is to use the protocol stacks IP/RFC 1490/AAL 5 (or AAL 3/4) and offer DS3 connections. Due to limited availability of ATM-related products, there exists an encapsulation mismatch problem with this architecture. In this article we present a (routing) solution to this problem using existing technologies. The key to this design is to overcome incompatibility problems by inserting a router and making it function like a layer 2 device via careful address assignment and the use of Proxy ARP. This design has proven to be a feasible solution during a joint ATM NAP Testing Project by Merit, Pac*Bell and Bellcore in May 1994.

Introduction

The new NSFNET architecture is composed of four components: *Network Access Points* (NAPs); a *Routing Arbiter* (RA), a *Very High-speed Backbone Network* (vBNS); and a set of *Regional Network Providers* (RNPs) as defined in [1]. They are expected to be in place by 4th quarter of 1994.

It is now commonly understood that a NAP is a (layer 2) infrastructure, to which a number of *Internet Service Providers* (ISPs) and other networks can connect via routers to exchange traffic. The *Route Server* (RS) attached to a NAP is to facilitate and simplify inter-domain routing exchange at the NAP by peering (via, e.g., external BGP [4]) with ISPs' routers and passing inter-domain routing information to these routers. A conceptual (logical) diagram of a NAP is shown in Figure 1 where ISPs' routers (R_1 and R_2) peer via BGP with the RS.

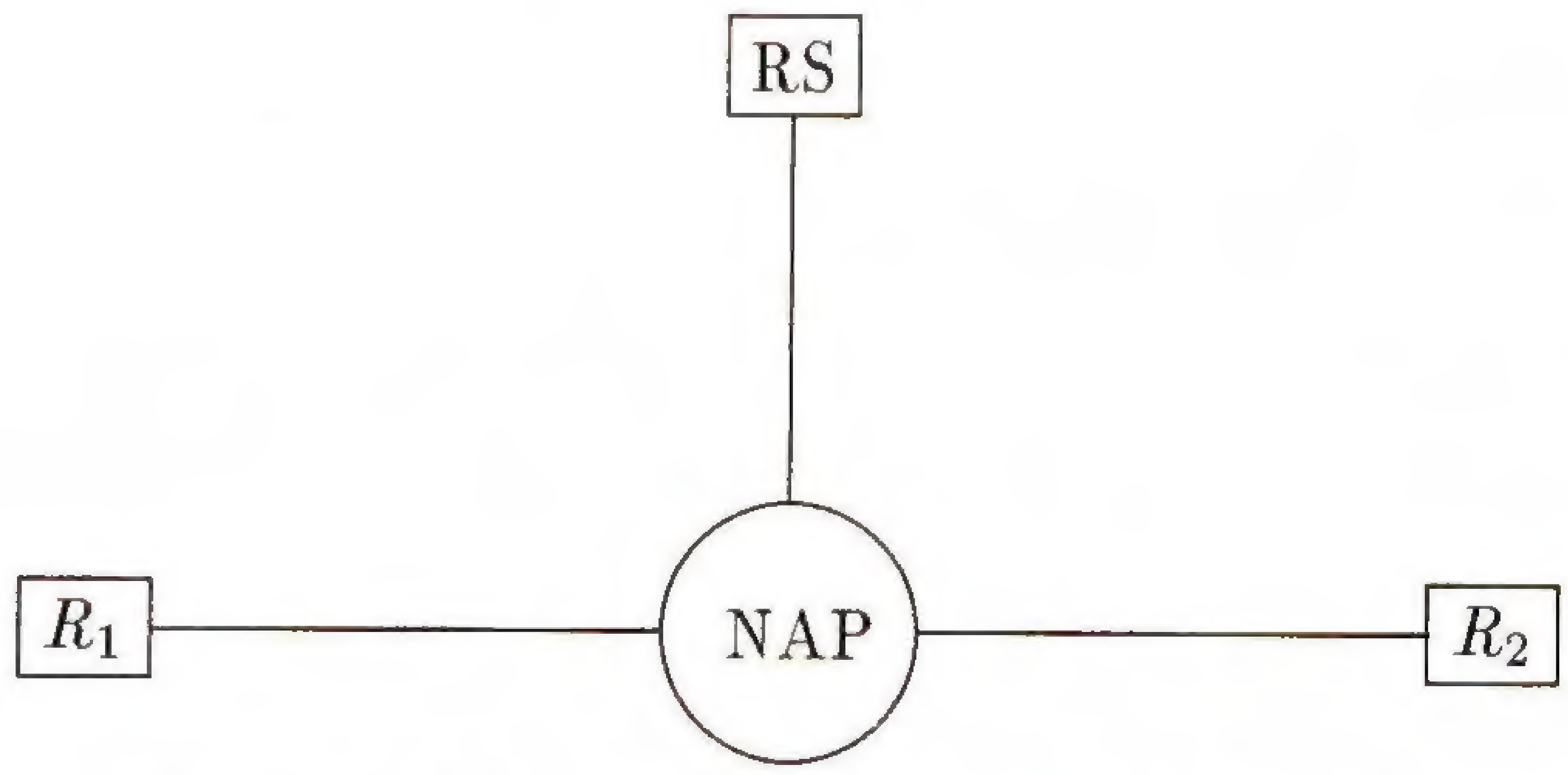

Figure 1: Conceptual Diagram of a NAP

It is noted that the RS itself is not involved in packet forwarding. To realize such functionality, the RS relies on the feature of the third-party routing information, available in routing protocols such as BGP (see Appendix C and [3] for more details). This feature requires that the RS and all the routers attached to a NAP be configured as a *Logical IP Subnetwork* (LIS), that is, they can reach each other in a single (logical) hop without resorting to routing. Before the actual deployment of NAPs, we must study the routing integrity issues, particularly how this one-hop requirement can be met in each of the proposed NAP sites.

The *Asynchronous Transfer Mode* (ATM) technology has been chosen by several NAP providers as the underlying NAP fabric [8]. Currently, due to lack of routers that support standard-based native IP over ATM encapsulation [5,7], and limited availability of ATM interfaces, routing in the ATM NAPs is not expected to be as straightforward as in some other media such as FDDI.

In this article we present a routing design, using currently-available technologies, to address the encapsulation mismatch problem with the initial ATM NAP architecture as currently planned. This design overcomes incompatibility problems by inserting a router and making it function like a layer 2 device via careful address assignment and the use of the Proxy ARP [9,10]. This design has proven to be a feasible solution during a joint ATM NAP Testing Project by Merit, Pac*Bell and Bellcore.

This article is organized as follows. The routing problem with the initial ATM NAP architecture is presented in the next section. This is followed by a summary of the routing design. Next we describe the setup of a testbed based on this routing design and present the routing testing results. The Appendix gives a brief review and clarification of several issues on ARP, Proxy ARP, and BGP Next Hop, which are all closely related to the presentation of this article. The Appendix also includes highlights of the testbed configuration.

Routing problem with initial ATM NAPs

Dictated by the availability of ATM-related products within the desired timeframe, the initial phase of the ATM NAPs planned by the NAP providers is as follows [8]:

- PVC-based architecture
- DS3 interface to ATM switch
- Frame Relay IP encapsulation (RFC 1490)

The Frame Relay IP encapsulation scheme is planned for the initial phase because the currently-available product (i.e., router) that connects to ATM with DS3 interface is based on RFC 1490. An ATM DSU (ADSU) is required in order to connect an ISP's router to the ATM switch. As a result, an IP packet is encapsulated into a Frame Relay packet and is then converted into ATM cells (AAL 3/4 or AAL 5) and passed to the ATM switch, as shown in Figure 2.

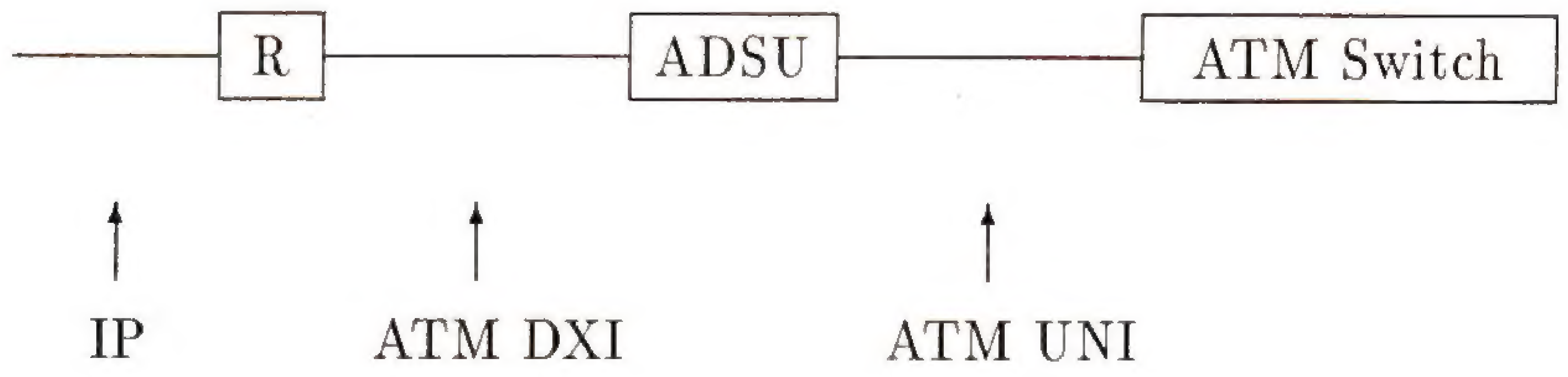

Figure 2: Current Connection to ATM

The Route Server attached to a NAP will be a Sun SPARC workstation running modified *GateD* software initially. Currently-available ATM interfaces cards for the Sun workstation support both NULL and LLC/SNAP encapsulation [5], yet the ATM DSU approach requires a NLPID encapsulation [6]. Even though both the ATM DSU and the workstation interface support AAL5, they are incompatible due to encapsulation mismatch of the IP packet in the AAL5 CS-PDU. Therefore, the RS cannot be directly connected to the ATM switch while the ISPs' routers are connected to the switch via the ATM DSUs. A straightforward solution to this incompatibility problem is to insert a router between the RS and the ATM switch. A logical diagram of the ATM NAP is shown in Figure 3, in which R_1 and R_2 are assumed to be ISPs' routers and R_0 is the router that connects the RS with the ATM switch.

continued on next page

Routing Design for the ATM NAPs *(continued)*

In this topology full-mesh PVCs need to be set up between all ISPs' routers that wish to communicate. The routers' interfaces attached to the ATM switch need to be configured as a LIS (but not necessarily with the same network prefix; see Appendix A for details), and their IP addresses need to be mapped to the data link connection identifiers (DLCIs) due to the point-to-point nature (virtual circuits) of the Frame Relay setup.

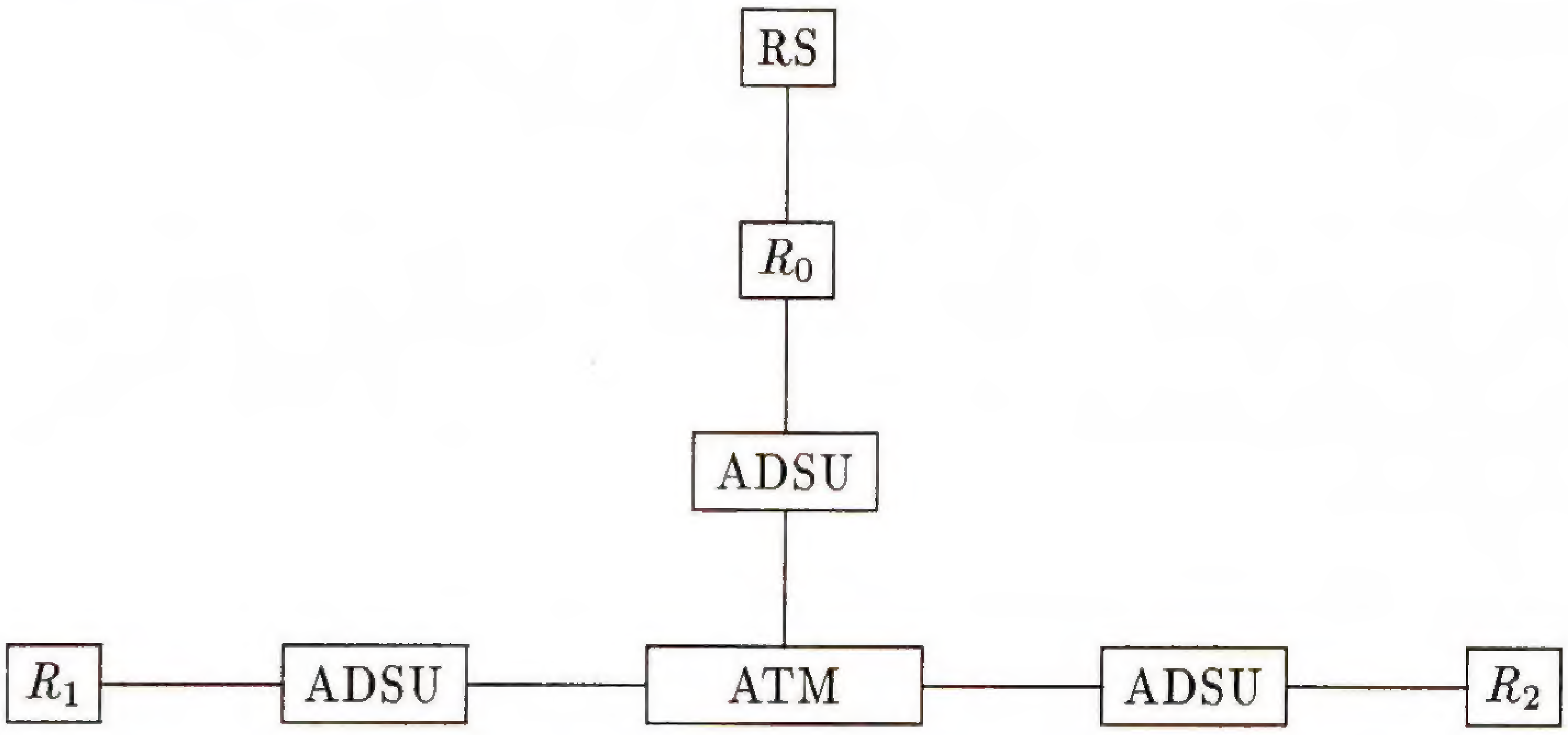

Figure 3: ATM NAP Structure

As was noted in the introduction, the concept of NAPs and Route Servers requires that the RS and the NAP-attached routers be able to reach each other in a single (logical) hop without relying on routing. Adding the router R_0, however, breaks this requisite. How to ensure routing integrity in this topology is the subject of the next section.

A routing solution

We have developed a solution, using currently-available technologies, to the routing problem presented in previously. The key to this solution is a scheme that makes the inserted router R_0 (of Figure 3) function as a layer 2 device so that the RS and the ISPs' routers can reach each other directly without routing. This design has several components:

(a) Careful address assignment so that the RS and the ISPs' routers appear to be in a LIS, meeting the requirement that the RS and the NAP-attached routers be one logical hop away from each other.

(b) Careful address assignment (see later under "Testing" and Appendix B for details) so that the router R_0 does Proxy ARP for packets sent from the RS to an ISP's router. (Note that a static route configuration on the RS can achieve the same goal here.)

(c) Configuration of a static mapping of the RS's address on the ISPs' routers (pointing to R_0) in order to pass packets from an ISP's router to the RS. (Note again that a static route configuration on the ISPs' routers can achieve the same goal here.)

(d) Utilization of the multi-hop-BGP-peer feature implemented in various routing software for peering between the RS and the ISP's routers.

When the Route Server sends a routing packet to an ISP's router, the Route Server will ARP for the destination since the router appears to be on the same subnet with the RS due to (a).

The router R_0 will return a Proxy ARP reply to the RS since R_0's other interface is on the same subnet with the destination router due to (b). As a result, the packet will be delivered from the RS to the router R_0 and forwarded to the destination router.

When an ISP's router sends a packet to the RS, it will use its statically-configured mapping to send the packet to R_0 due to (c). R_0 will deliver the packet to the RS.

The RS peers via BGP with the ISPs' routers and passes routes to them based on their routing policy. Since the RS and the ISPs' routers appear to be on the same subnet, the routers will receive routes pointing to the correct next hop (i.e., another ISP's router, not the RS). Therefore, the RS is not involved in packet forwarding. (See Appendix C for details.)

Testing

The above design has been tested during a joint ATM NAP Testing Project by Merit, Pac*Bell and Bellcore. The testbed setup and testing results are presented in this section. This testbed was provided by Pac*Bell.

In this particular test, the ADSUs used support AAL 3/4. The current plan [13] is to use ADSUs that support AAL 5 (e.g., ADSUs by ADC Kentrox). However, the routing design presented in the article is still suitable.

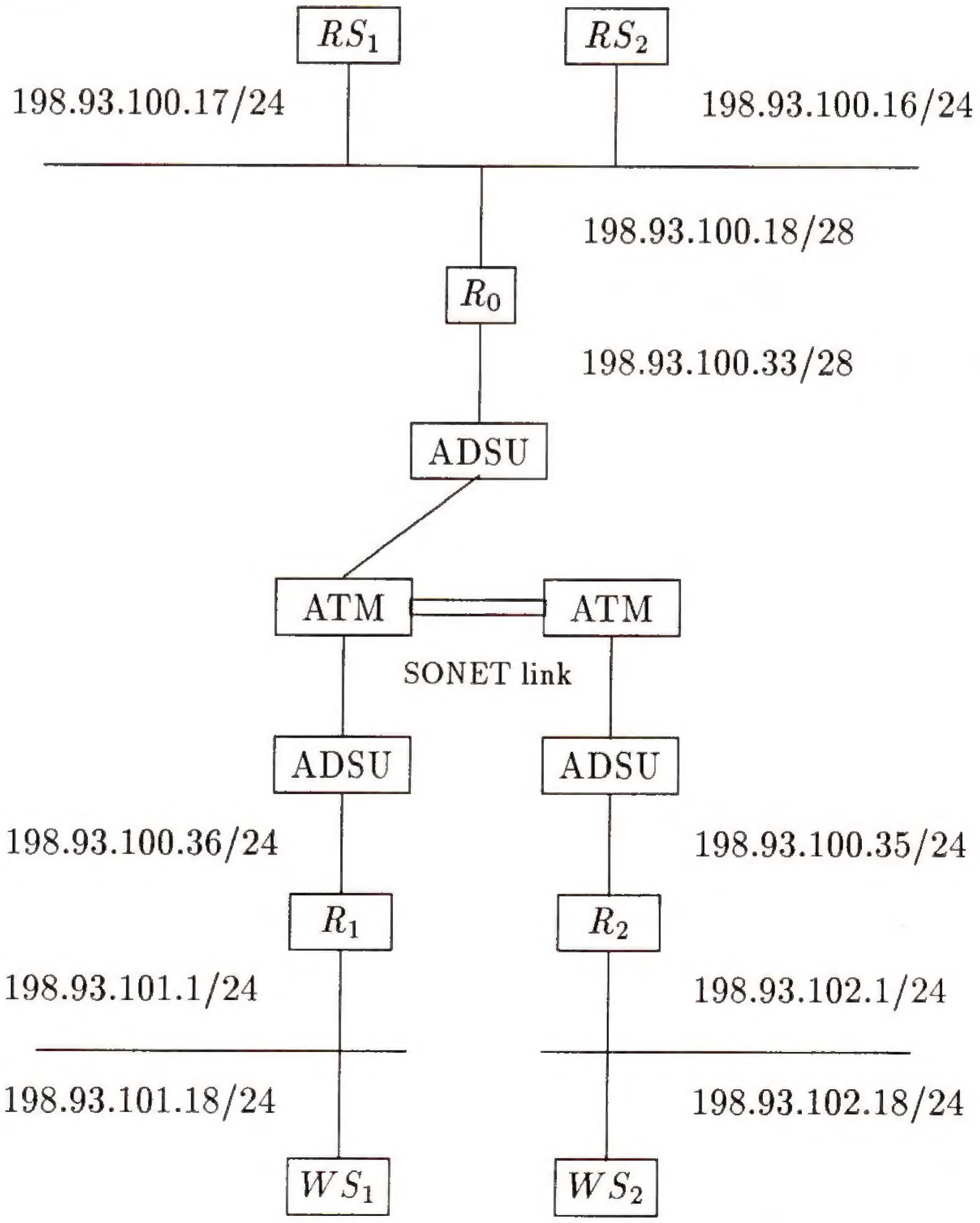

Figure 4: Testbed Topology Map
(R_1 and R_2 are ISPs' routers and R_0 connects the RSs with the ATM switch)

continued on next page

Routing Design for the ATM NAPs *(continued)*

Equipment

The following is a list of major equipment that is used to set up the testbed:

- Newbridge ATM switch(es)
- SPARC-10 workstations as Route Servers, with *GateD* Release 3.5 alpha 4
- Cisco 7000 routers as ISPs' routers, with software version 10.0 (0.17)
- Cisco AGS+ router as the one between the RS and the ATM switch
- ADSUs (DL3200A by Digital Link) to connect Cisco routers to the ATM switch

Topology

The testbed topology is shown in Figure 4. In this setup, the two ATM switches are connected together by a SONET link to form the NAP fabric. RS_1 and RS_2 are connected with R_0 via an Ethernet and they are assumed to be the primary and secondary RSs, respectively, for the purpose of redundancy. Routers R_1 and R_2 are assumed to be different ISPs' routers attached to the NAP. Workstations WS_1 and WS_2 are connected via Ethernet with R_1 and R_2, respectively. They are assumed to belong to different ISPs' domains and are used to generate traffic and test connectivity.

In this article we use the notation "x/p" to mean that the IP address of the interface is x and the network prefix is the first p high-order bits of x (in other words, the first p high-order bits of the network mask are all ones, and the rest are all zeros).

Configuration

The configuration highlights for the RS, R_0, and the ISPs' routers (R_1 and R_2) are listed in Appendix D. Note that:

- The interfaces (attached to the NAP) of RS_1, RS_2, R_1 and R_2 are configured as in a LIS, and R_0 is configured with longer network prefixes. (See Appendix B for details.)
- BGP peering sessions exist between the RSs and the ISPs' routers. The RSs receive the route `198.93.101.0` from R_1 and pass it to R_2. Similarly, the RSs receive the route `198.93.102.0` from R_2 and pass it to R_1.
- There is no peering session between the ISPs' routers (R_1 and R_2).

Testing results

The following list highlights the testing results:

- As expected, routing information is exchanged between the RS and the ISPs' routers.
- More importantly, each ISP's router has routes with the desired next hop information, i.e., the RS is not involved in traffic forwarding. For example, in R_2's routing table the next hop for network `198.93.101.0` is R_1.
- The two RSs backup each other seamlessly, as expected.
- The testing showed that the concept of Route Servers and ATM NAPs works with today's technology.

Discussion

The routing design presented addresses the encapsulation mismatch problem with the initial ATM NAP architecture as currently planned. The ATM NAP architecture itself, including the selection of the protocol stacks, is briefly reviewed but it is not the subject of the article.

It is fully anticipated that the ATM NAP architecture will evolve as the ATM technology advances and more ATM-related products become available. However, this design could still be useful in the future architecture, for instance, if the RS wishes to connect to the ATM with DS3 bit rate but only OC3 or TAXI interfaces are available for the RS. In general, we believe that the design presented in this article, perhaps with small variations, can be applied in situations when the physical topology of a network needs to be hidden from the logical network layer. For example, it can be applied to address interface mismatch problems in other types of NAPs (e.g., SMDS), or when an ISP's routers could not accommodate an interface matching a particular NAP medium. In these cases, a router with compatible interfaces could be used to connect an ISP's router with the NAP, as illustrated in this article. Note that the router in between has only one forwarding table. So, if multiple ISPs want to share that router, their routing policy must be identical and that router could become a bottleneck for traffic throughput.

The shortcomings of this design include:

- Cost of one additional router;
- Increase in failure probability associated with it;
- An extra hop (in reality) for routing traffic;
- Some increase in management complexity (static configuration).

Therefore, this design is not recommended unless the situation warrants its use.

Acknowledgement

We would like to thank Pac*Bell and Bellcore, especially Frank Liu, Chin Yuan, Mike Rudik, Liang Wu and Al Broscius for providing the testbed equipment and for their contribution to the joint ATM NAP Testing Project. We also acknowledge useful comments on an earlier version of this document from members of the Internet Engineering Group of Merit.

In the following Appendices we give a brief review of ARP, Proxy ARP, and BGP Next Hop, which play essential roles in the design presented in this article. We first explain the general principles, and then illustrate, through examples (when appropriate), the technical details that are most relevant to the discussions in this article.

Appendix A: Address Resolution Protocol (ARP)

The *Address Resolution Protocol* (ARP) is used to perform dynamic address translation between (broadcast) LAN hardware addresses and Internet addresses. It was original designed in [9] for Ethernet and has since been extended to operate over many non-Ethernet LANs, including SMDS [11] and FDDI [12].

A host (including router) attaches to a LAN via an interface. The interface is usually configured with both an Internet address, and a network prefix (or mask) which determines the network the interface is directly attached to.

continued on next page

Address Resolution Protocol *(continued)*

A host (say *S*) uses ARP to find the MAC layer address of a destination host *D* (identified by an IP address) by broadcasting an ARP request. Host *D* responds to the ARP request by sending a reply that contains its MAC layer address. Then host *S* sends packets to host *D*'s MAC layer address.

- *When does a host use ARP?:* When host *S* wants to send a packet to host *D,* host *S* first checks if host *D*'s address falls under host *S*'s network prefix. If "yes," host *S* would think that host *D* is also on the directly attached network specified by host *S*'s network prefix. Thus, host *S* resorts to ARP and then send the packet to host *D*'s MAC layer address directly, no routing is involved; if "no," host *S* resorts to routing.
- *Directly reachable hosts are not required to have the same network prefix:* Clearly, two hosts are directly reachable (via ARP) if each host falls under the other's network prefix, i.e., they share the longer network prefix. In such case, we also say that these hosts are in a "Logical IP Subnet" (LIS). For example, hosts that have Internet addresses `198.93.100.17` with mask `255.255.255.0` and `198.93.100.18` with network mask `255.255.255.240`, respectively, can reach each other directly via ARP as they share the longer network prefix `198.93.100.18/28`.

Appendix B: Proxy ARP

Proxy ARP was proposed in [10] and has been implemented in virtually all routers. It allows one network address to be shared between two (broadcast-type) physical nets connected by a gateway. The gateway sends an ARP reply specifying its MAC layer address in response to an ARP request for a target IP address which is not on the directly-connected network but for which the gateway offers an appropriate route.

The ATM NAP Testing Project uses a special setup that we call "symmetric Proxy ARP." In the following we summarize the structure and address assignment strategy for such a setup.

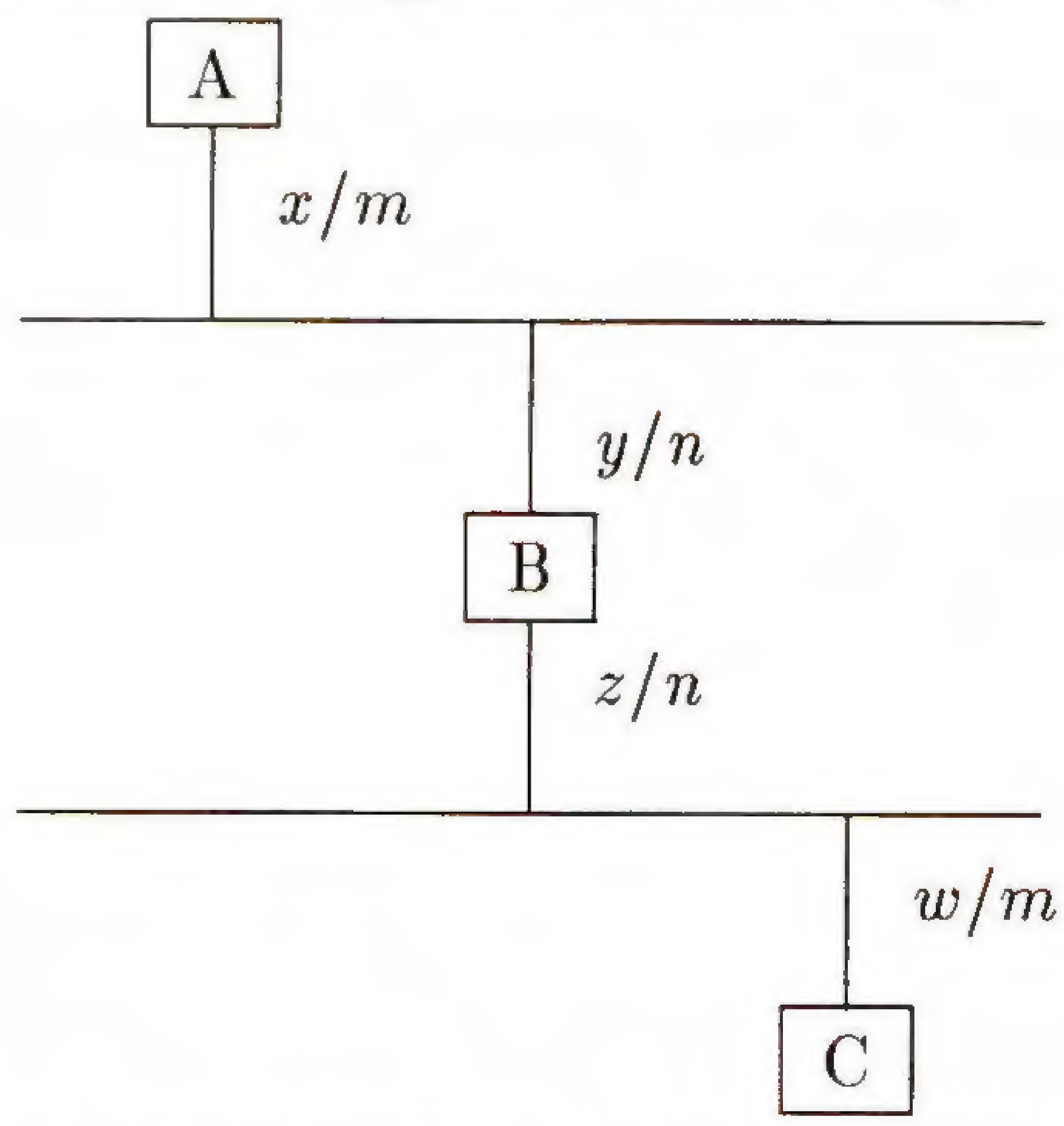

Figure 5: Symmetric Proxy ARP

In the setup as shown in Figure 5, hosts A and C have the same network prefix length and the two interfaces of B also have the same network prefix length. We call such setup "Symmetric Proxy ARP" if hosts A, B, C can reach each other directly (via ARP and Proxy ARP).

This full reachability (via ARP and Proxy ARP) would mean: (for convenience an interface is referred to by its configuration.)

x/m and y/n:	a LIS
x/m and w/m:	a LIS
w/m and z/n:	a LIS
y/n and z/n:	different subnets.

It is easy to verify that the above statements translate into the following formal conditions for the "Symmetric Proxy ARP":

- $m < n$
- y and z share m, but less than n, high-order bits
- x and y share n high-order bits
- z and w share n high-order bits.

Based on these conditions, the address assignment strategy would be: first assign network prefix length m, n (integers); then assign IP addresses y, z; then assign IP addresses x, w.

Appendix C: BGP Next Hop

The Route Server exchanges inter-domain routing information without being involved in packet forwarding. It relies on the feature of third-party routes in routing protocols such as BGP. In this appendix we explain the "third-party BGP" through examples.

The following is an excerpt from [4] (Sect. 5.1.3), which describes the so-called "third party BGP":

> "The NEXT_HOP path attribute defines the IP address of the border router that should be used as the next hop to the networks listed in the UPDATE message. ... A BGP speaker can advertise any external border router as the next hop, provided that the IP address of this border router was learned from one of the BGP speaker's peers, and the interface associated with the IP address of this border router (as specified in the NEXT_HOP path attribute) shares a common subnet with the local and remote BGP speakers."

The "third party BGP" is explained in the following examples.

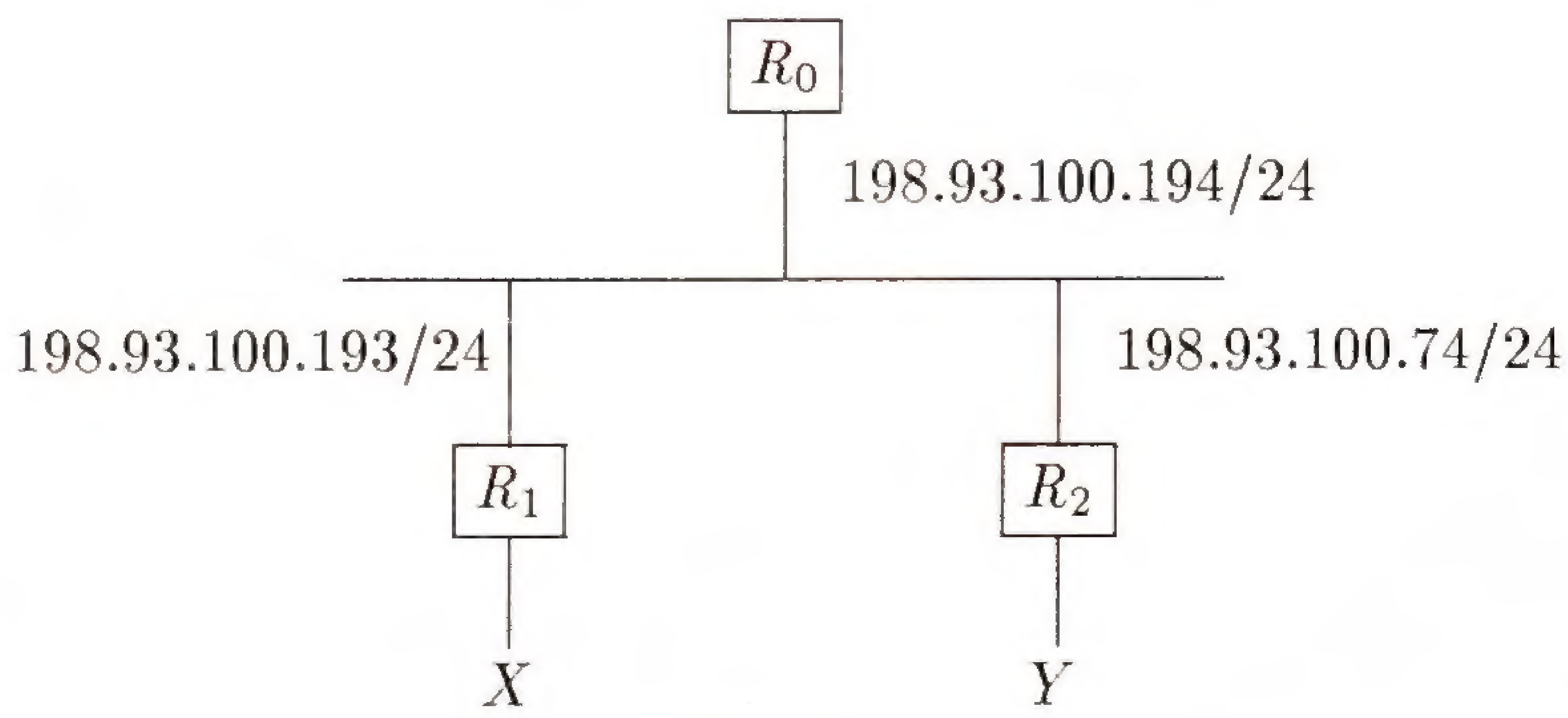

Figure 6: For Example 1

Example 1: In Figure 6, R_0 peers via (external) BGP with R_1 and R_2. R_1 and R_2 have routes to network X and Y, respectively. There is no peering session between R_1 and R_2.

continued on next page

BGP Next Hop *(continued)*

Due to the BGP sessions, R_0's routing table would have Y with R_2 as the next hop. Because R_1 and R_2 fall under R_0's network prefix, R_0 would pass Y to R_1 with R_2 as the next hop. R_1 would then install this route without modification to the NEXT_HOP because R_2 falls under R_1's network prefix (i.e., R_1 can directly reach R_2). By the same token, R_2 would have X with R_1 as the next hop.

In such a case, R_0 only passes routing information and is not involved in packet forwarding. Functionally R_0 can be viewed as a router server (as defined for a NAP).

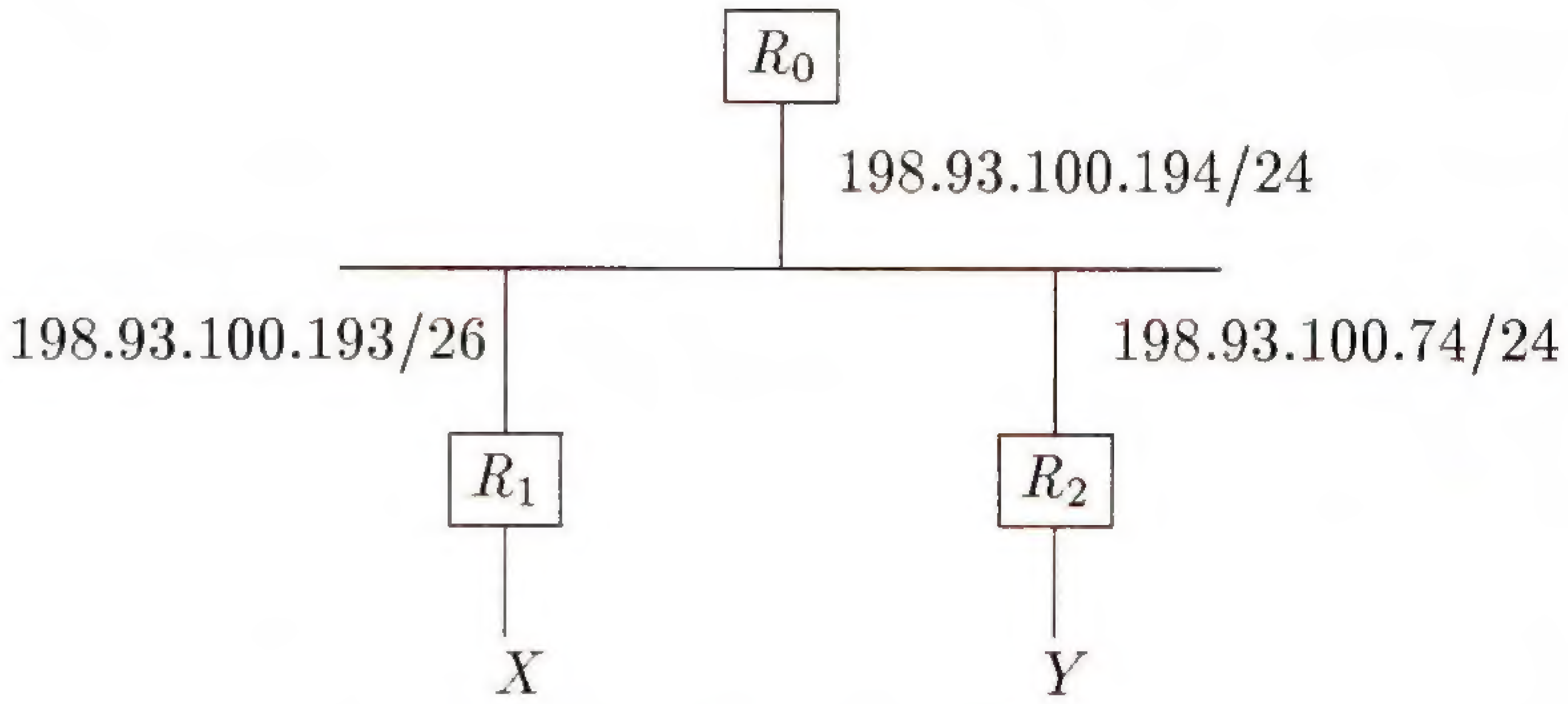

Figure 7: For Example 2

Example 2: Only the address of R_1 (Figure 7) differs with Example 1. However, R_1 can not directly (via ARP) reach R_2 as R_2 does not fall under R_1's network prefix.

In this example, R_0 would still pass Y to R_1 with R_2 as the next hop. However, R_1 would not accept (from R_0) Y with R_2 as the next hop because R_2 does not fall under R_1's network prefix (i.e., R_2 is not directly reachable). Instead, R_1 would install R_0 as the next hop for Y because Y is announced by R_0 and there must exist a route from R_0 to Y. So packets from R_1 to Y would go through R_1 —> R_0 —> R_2 —> Y. Also, it is easy to verify that packets from R_2 to X would still go through R_2 —> R_1 —> X.

Appendix D: Testbed Configuration Highlights

Router R_0:

```
!
interface Ethernet1
ip address 198.93.100.18 255.255.255.240
!
interface Hssi0
ip address 198.93.100.33 255.255.255.240
encapsulation frame-relay
no keepalive
frame-relay map ip 198.93.100.35 106 broadcast
frame-relay map ip 198.93.100.36 104 broadcast
!
```

Router R_1 (Cisco 10.0):

```
!
interface Ethernet0
ip address 198.93.101.1 255.255.255.0
!
interface Hssi1
ip address 198.93.100.36 255.255.255.0
encapsulation frame-relay
no keepalive
frame-relay map ip 198.93.100.17 107
frame-relay map ip 198.93.100.33 107 broadcast
frame-relay map ip 198.93.100.35 103 broadcast
!
router bgp 185
network 198.93.100.0
network 198.93.101.0
neighbor 198.93.100.17 remote-as 183
neighbor 198.93.100.17 ebgp-multihop
!
```

Router R_2 (Cisco 10.0):

```
!
interface Ethernet0
ip address 198.93.102.1 255.255.255.0
!
interface Hssi2
ip address 198.93.100.35 255.255.255.0
encapsulation frame-relay
no keepalive
frame-relay map ip 198.93.100.17 105
frame-relay map ip 198.93.100.33 105 broadcast
frame-relay map ip 198.93.100.36 102 broadcast
!
!
router bgp 184
network 198.93.100.0
network 198.93.102.0
redistribute static
neighbor 198.93.100.17 remote-as 183
neighbor 198.93.100.17 ebgp-multihop
!
```

RS_1 (GateD 3.5):

```
interfaces {
interface 198.93.100.17 as 183;
};
autonomoussystem 183;
routerid 198.93.100.17;
rip no;

bgp yes {
preference 50;
group type external peeras 184
{
peer 198.93.100.35
passive
ttl 3;
};
```

continued on next page

Testbed Configuration Highlights *(continued)*

```
group type external peeras 185
{
peer 198.93.100.36
passive
ttl 3;
};
};

import proto bgp as 184 {
198.93.102 masklen 24 preference 101;
all;
};

import proto bgp as 185 {
198.93.101 masklen 24 preference 101;
all;
};

export proto bgp as 184 {
proto bgp as 185 {
198.93.101 masklen 24 exact;
};
};

export proto bgp as 185 {
proto bgp as 184 {
198.93.102 masklen 24 exact;
};
};
```

References

[1] "Network Access Point Manager, Routing Arbiter, Regional Network Providers, and Very High Speed Network Services Provider for NSFNET and NREN(SM) Program," Program Solicitation, National Science Fundation, May 1993.

[2] Estrin, D., Postel, J., and Rekhter, Y., "Routing Arbiter Architecture," *ConneXions,* Volume 8, No. 8, August 1994.

[3] Ford, P., "NAPs Do Not Route," draft, April 1994.

[4] Rekhter, Y., and Li, T. (Editors), "A Border Gateway Protocol 4 (BGP-4)," RFC 1654, July 1994.

[5] Heinänen, J., "Multiprotocol Encapsulation over ATM Adaptation Layer 5," RFC 1483, July 1993.

[6] Bradley, T., Brown, C., and Malis, A., "Multiprotocol Interconnect over Frame Relay," RFC 1490, July 1993.

[7] Laubach, M., "Classical IP and ARP over ATM," RFC 1577, January 1994.

[8] Broscius, A., "Overview of Joint NAP Proposal: AADS, Bellcore, and Pacific Bell," presentaion at the First North America Network Operators Group Meeting (NANOG), Ann Arbor, June 1994. Slides available from Internet host `merit.edu` in directory `ftp/pub/nanog/presentations/`.

[9] Plummer, D., "An Ethernet Address Resolution Protocol," RFC 826, November 1982.

[10] Postel, J., "Multi-LAN Address Resolution," RFC 925, October 1984.

[11] Piscitello, D. and Lawrence, J., "The Transmission of IP Datagrams over the SMDS Service," RFC 1209, March 1991.

[12] Katz, D., "A Proposed Standard for the Transmission of IP Datagrams over FDDI Networks," RFC 1188, October 1990.

[13] Salo, T.J., "ATM NAP Workshop Conclusions and Summary," Draft, September 1994.

[14] Laubach, Mark, "ATM for your internet—But When?" *ConneXions,* Volume 7, No. 9, September, 1993.

[15] ATM Forum, "User Network Interface (UNI) Specification Version 3.0," ISBN 0-13-225863-3, Prentice Hall, December 1993.

[16] Atkinson, Ran, "Towards Real ATM Interoperability," *ConneXions,* Volume 7, No. 8, August 1993.

[17] Laubach, Mark, "IP Over ATM and the Construction of High-Speed Subnet Backbones," *ConneXions,* Volume 8, No. 7, July, 1994.

[18] Malis, A., "Multiprotocol Encapsulation over Frame Relay," *ConneXions,* Volume 6, No. 8, August 1992.

[19] Kozel, E., "The Cisco/DEC/NTI/StrataCom Frame Relay Specification," *ConneXions,* Volume 5, No. 3, March 1991.

JESSICA (JIE YUN) YU received her B.S. in Computer Science from Shanghai Jiao Tong University in China and M.S. in Computer Engineering from The University of Michigan. She is currently a Senior Internet Engineer of Merit Network Inc. and is one of the co-PIs of the Routing Arbiter Award from NSF. She joined Merit in 1987 and was a key contributor in the initial engineering and deployment of the T1 and T3 NSFNET backbone and has since worked on the NSFNET engineering and operation. The major focus of her work has been NSFNET routing architecture design, engineering and implementation, and Internet advanced routing research. She has performed pioneering working on BGP and its deployment and is one of the architectors of CIDR. She co-authored several RFCs and co-chairs the IETF CIDR Deployment Working Group. She is currently working on various Routing Arbiter related tasks including Routing Registry design, NAP routing design and ATM-NAP testing. She is also a member of FEPG (Federal Networking Engineering Planning Group). E-mail: `jyy@merit.edu`

ENKE CHEN received the B.S. degree in Instrumentation and Automation from Tsinghus University, Beijing, China, in 1985, and the M.S.E. and Ph.D. degrees in Electrical Engineering: Systems from the University of Michigan, Ann Arbor, in 1987 and 1991, respectively. Mr. Chen joined Merit as an Internet Engineer in 1992. He has worked on the logical modeling and performance analysis of discrete-event dynamic systems, NSFNET routing design and engineering, developing software to study and improve the NSFNET routing and its configuration process, and more recently on ATM-NAP testing and routing design, and other Routing Arbiter related projects. Mr. Chen has authored several journal papers including IEEE Transactions. He is a member of IEEE since 1988. E-mail: `enke@merit.edu`

LAURENT JONCHERAY received a B.S degree in Electronic from the Ecole Centrale de Paris in 1990 and a M.S. degree in Computer Networks from the Telecom Paris in 1991. He worked one year at CERN (Switzerland) in the Network Department and joined Merit Network, Inc. in 1993. He has developed tools to automate router configuration generation and built software for network traffic statistic collection and analysis. He is currently an Internet Engineer involved in the design and implementation of various Routing Arbiter tasks. His work focuses on router configuration tools, routing analysis, security and ATM-NAP testing. E-mail: `lpj@merit.edu`

On Traffic Measurements that Defy Traffic Models (and vice versa):
Self-Similar Traffic Modeling for High-Speed Networks

by
Walter Willinger, Daniel V. Wilson & Will E. Leland, Bellcore
and Murad S. Taqqu, Boston University

Introduction

In the past years, the networking and data communications community has become noticeably frustrated with proposed traffic models for today's packet networks. The general dissatisfaction stems from the facts that the proposed models (1) have little in common with network engineers' practical experience about the nature of the traffic observed in their networks, (2) are almost never validated using actual traffic measurements, and (3) have become increasingly complex without providing real insight into packet network traffic dynamics.

Recent findings from detailed studies of actual traffic measurements from a variety of different working packet networks have shown that this frustration is fully justified: the traditional traffic models have little to do with reality, and it is surprisingly easy to clearly distinguish empirically between actual network traffic and traffic generated by these currently used models. In this article, we give reasons for the present collision course between network engineers and traffic modelers and summarize the results of the recent studies of large sets of actual network traffic data. We also comment on how these results motivate and influence the development of *self-similar* or *fractal* models for high-speed network traffic, and we present preliminary evidence for how these new models impact practically all aspects of network engineering. However, a significant research effort is needed in the near future to use these new models for the purpose of building tomorrow's first-rate high-speed networks.

A Folklore Theorem

Who has not been told at some point during his or her math or engineering upbringing about the *folklore theorem* in teletraffic theory that says "multiplexing a large number of independent traffic streams results in a Poisson (i.e., smooth) process?" So common is this folklore result that traffic on the main communications links in today's packet networks is commonly claimed to be Poisson or Poisson-like. How could it be not! Apparently, even ATM switch vendors have been sold on this popular belief and have produced first-generation ATM switches with small buffers (between 10–100 cells), exactly as recommended by the traffic engineers who based their buffer sizing decisions on the Poisson nature of the expected traffic on the input links to the switch. However, recent articles in the trade press [1] made it clear that something must have gone terribly wrong; when deploying these switches in the field and exposing them to *real* traffic, cell losses far beyond normal were experienced and resulted in a redesign of the switches. What went wrong? Why did real traffic not behave according to theory? Clearly, there is nothing wrong with *Palm's Theorem* [2] which provides the mathematical underpinning for the folklore theorem. Thus, the problem must be with the individual traffic streams that are multiplexed and the fact that they violate some of the assumptions under which Palm's Theorem holds. But what are the traffic characteristics of the individual packet streams that seem to invalidate the folklore theorem, and what should we expect actual traffic transported on the main links in tomorrow's networks to look like, if not Poisson?

From POTS to B-ISDN

Historically, traffic modeling has been associated with *Plain Old Telephone Service* (POTS), and has been based almost exclusively on *Poisson* (or more generally, Markovian) assumptions about traffic arrival patterns and on *exponential* assumptions about resource holding requirements. It can be argued that traditional traffic modeling for telephone networks, in combination with the corresponding queueing and performance analysis, has been one of the most successful applications of mathematical modeling in industry. It has lead to today's first-rate telephone networks whose *Quality-Of-Service* (QOS) we fully rely on and take for granted.

However, POTS networks are of the past, and the future belongs to the *Broadband Integrated Services Digital Network* (B-ISDN) based on the *Asynchronous Transfer Mode* (ATM) technology, better known under the well-publicized name of the "National Information Superhighway," B-ISDN promises to be a single, high-speed, service-independent, flexible and efficient network that transports voice, data and video. ATM offers the speed and performance required for this purpose; it is supposed to handle aggregate traffic streams of different QOS; and it is expected to provide essentially infinitely scalable connectivity. Of course, it will take some time before the planned National Information Superhighway becomes a reality. In the meantime, the challenge for today's traffic engineers is to gain an understanding of the likely nature of the future B-ISDN traffic and to use this knowledge when building B-ISDNs that are expected to be of similar quality and reliability to today's telephone networks.

From a traffic modeling perspective, the emergence of modern high-speed packet networks (e.g., LANs, Gigabit networks, experimental ATM networks) and the prevailing trend toward B-ISDNs combines drastically new and different transmission and switching technologies with dramatically heterogeneous mixtures of services and applications. As a result, the use of traditional (i.e., telephony-based) traffic processes for modeling traffic in modern high-speed communications systems has come under intense scrutiny, especially because of the complete absence of validations of these models against measured network traffic. Their use is further questioned by the apparent differences that exist between the behavior of traffic experienced on actual networks and the traffic predicted by these models.

The traditional approach

Everybody agrees that packet traffic is more complex or *bursty* than voice traffic, simply because it is spanning vastly different time scales, from microseconds to seconds and minutes. However, beyond this point, it is hard to imagine more disagreement, and there seem to be as many definitions of *burstiness,* ways to measure burstiness and attempts to model burstiness as there are researchers trying to develop accurate traffic models for today's packet networks. Rooted firmly in the classical Poisson paradigm, traditional traffic modeling has approached the problem of how to handle burstiness in an incremental fashion, moving over time from batch-Poisson and interrupted Poisson models to 2-state or higher-state Markov Modulated Poisson Processes (MMPPs), all the way to the versatile Markovian Arrival Processes (MAPs) of Neuts [3] and the Batch-Markovian Arrival Processes (BMAPs) introduced by Lucantoni [4].

The hallmarks of this approach to traffic modeling and of the resulting traffic models are:

- The modeling is mainly driven by the desire of maintaining the analytic tractability of related queueing and performance problems

continued on next page

Self-Similar Traffic Modeling *(continued)*

- The adequacy of a model is almost never judged by how well it fits actual traffic data in a statistical sense but by how well it predicts performance of a given queueing model
- A proposed model is "experimentally verified" by comparing a Monte Carlo simulation of the model's (!) traffic with the analytical results
- The approach typically requires a large number of model parameters, and practical problems related to their estimation and interpretation, and to the investigation of critical parameters that drive pragmatic concerns of network design and operations are only mentioned (at best) in passing.

Given this list, it is not surprising to observe these days a generally very negative reaction by network and traffic engineers about the state-of-the-art of traffic modeling and performance analysis for today's packet networks. In their view, the guidelines and tools they receive from the research community for their network engineering problems are often only of "academic interest" and of no or only little practical relevance: how can one trust engineering guidelines that are based on traffic models that have little in common with one's practical experience about the actual behavior or *real* network traffic?

The new "Data Driven" approach

One of the main reasons for this "estrangement" between network engineers on one side and traffic modelers on the other side has been the unavailability of actual traffic measurements from working packet networks. In fact, a survey by Pawlita [5] notes that between 1966 and 1987, several thousand papers on queueing problems have been published, but only about 50 on traffic measurement results! However, more recently, increasing volumes of traffic measurements from working networks (e.g., CCSN/SS7 at 56 Kbps, ISDN at 1.5 Mbps, Ethernet LANs at 10 Mbps) have been collected and made available to researchers. In view of the results reported below, it is safe to expect that in the near future, there will be no shortage of empirical traffic data from tomorrow's high-speed networks. As a result, *data-driven* traffic modeling approaches become feasible and promise to bridge the gap between a network practitioner's experience about the actual nature of network traffic and the currently proposed mathematical models thereof. Here, by "data-driven" we mean:

- The modeling is driven by the desire to capture and describe the essential characteristics uncovered by a rigorous statistical analysis of the traffic data at hand
- The adequacy of a proposed model is judged by how well it fits the actual traffic data in a statistical sense
- Experimental verification of a proposed model consists of comparing a Monte Carlo Simulation using the actual traffic data (trace-driven simulation) with that using the model's traffic and (if available) with the analytical results
- The modeling is *parsimonious,* resulting in a small number of model parameters (*traffic characteristics*) that can be (i) estimated from actual traffic measurements, (ii) given a physical interpretation, and (iii) shown to have a dominant impact on performance.

Clearly, this data-driven approach to traffic modeling relies heavily on advanced statistical techniques and up-to-date data analysis tools.

At the same time, the area of high-speed network traffic is a source for data sets that are unique in size and quality and allow for statistical analyses that would have been unheard of, even 5 years ago. As a result, it should come as no surprise that this data-driven modeling approach results in traffic processes that are often in stark contrast to those obtained through the traditional approach described earlier and seem to describe "reality" more accurately. After all, the distinguishing feature of the data-driven approach is the discovery of the basic characteristics of actual network traffic through statistical analyses, while the traditional method relies on certain preconceived opinions about network traffic (e.g., Markovian nature) that have not been checked against empirical data but are imposed for mathematical convenience.

Ideal example: Ethernet LAN Traffic

An ideal test case for modeling approaches for high-speed network traffic is Ethernet LAN traffic: it is of practical importance, and enormous data sets of Ethernet traffic measurements are available. The practical importance of LAN traffic such as Ethernet traffic stems from today's already widespread use of LANs, its realistic mixtures of use from established communities (e.g., industry, universities, government), and its importance as one of the expected major source (i.e., LAN interconnection service) for future B-ISDN traffic. Thus, understanding the basic characteristics of LAN traffic will provide valuable insight into the nature of realistic future B-ISDN and other high-speed network traffic scenarios. In turn, this insight can be expected to guide the future development of realistic and practically relevant traffic models.

Observations

Our statistical analysis of many hours worth of measured Ethernet LAN traffic collected at Bellcore from 1989–1993 is described in detail in [6,7,8]. Here, we briefly summarize the main findings and discuss their impact on traffic modeling for future high-speed networks. The main observations from our studies of actual Ethernet LAN traffic data are the following:

• *Self-Similarity:* Aggregate Ethernet LAN traffic (i.e., number of bytes or packets per time interval) looks the same regardless of the choice of time interval ("time scale") over which we observe the traffic process. In other words, if we omit the axis labels on plots displaying the number of of bytes or packets per time interval vs. time, for different choices of time intervals, ranging from milliseconds to seconds and minutes, it is nearly impossible to associate the different plots with the correct time scale. This behavior is illustrated in figure 1 on the next page with a plot sequence of time series of packet counts for 5 different time scales. Starting with a time unit of 100 seconds in Plot (a), each subsequent plot is obtained from the previous one by increasing the time resolution (i.e., decreasing the time unit) by a factor of 10 and then focusing on a randomly chosen subinterval (indicated by a darker shade in each plot). The time unit corresponding to the finest time scale is 10 milliseconds in Plot (e). Observe that all plots are visually very "similar" to each other, so that packet arrival rates measured over large time scales (hours, minutes) are practically indistinguishable from those measured over small time scales (seconds, milliseconds). In particular, within the time scales of interest, there is no natural length of a "burst": at time scales ranging from milliseconds to minutes and hours, traffic bursts have the same qualitative appearance. Processes with this property are called *(exactly) self-similar* or, using a more popular term, *fractal.*

continued on next page

Self-Similar Traffic Modeling *(continued)*

This observed self-similar behavior of measured Ethernet LAN traffic is immediately and visibly distinct from traffic generated by currently used theoretical models of packet traffic. The latter typically gives rise to plots of packet counts which quickly become indistinguishable from *white noise* (i.e., a sequence of independent and identically distributed packet counts) when plotted over increasingly larger time intervals, as illustrated by the plot sequence (a')–(e') in Figure 1. This sequence was obtained by successive aggregations as in the empirical plot sequence (a)–(e), except that it arose from synthetic traffic generated from a comparable compound Poisson process with the same average packet size and arrival rate as the empirical data. While the choice of a compound Poisson process is admittedly not very sophisticated, more complex Markovian arrival processes were observed to give rise to plot sequences indistinguishable from (a')–(e'). In this sense, the most striking result of our analysis is that using simple plot sequences derived from the traffic data, one can clearly distinguish the measured Ethernet LAN traffic from traffic generated by practically all traditional models for packet traffic currently proposed in the literature.

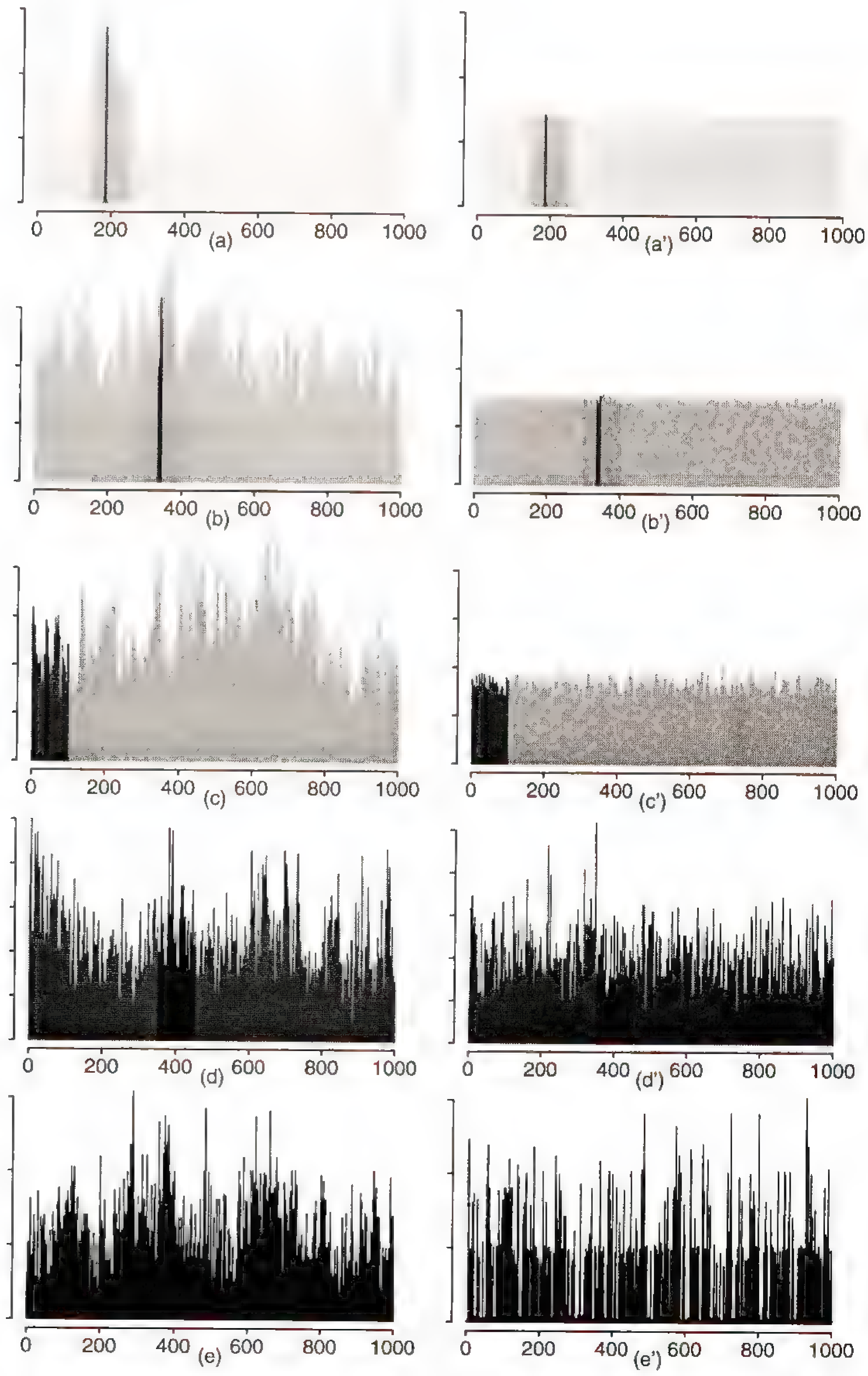

Figure 1: Self-similar Ethernet traffic (a)–(e), and synthetic traffic generated from a traditional model (a')–(e').

This fact directly challenges the use of these models for network and traffic engineering purposes. Moreover, it fully agrees with network engineers' criticisms about a total disagreement between the traffic behavior predicted by currently considered theoretical models and their practical experience with actual network traffic behavior.

- *Long-Range Dependence or The Joseph Effect:* Mathematically, the self-similarity property of Ethernet LAN traffic manifests itself in the presence of *long-range dependence;* traditional traffic models display a common feature called *short-range dependence.* Long-range dependence or, using Mandelbrot's terminology, *The Joseph Effect* (in reference to the Biblical figure who foretold of the "seven fat years and 7 lean years" that ancient Egypt was to experience), captures the persistence phenomenon observed in many empirical time series that manifests itself in clusters ("bursts") of consecutive large or small values. Slightly more formally, long-range dependent traffic processes (i.e, number of bytes/cells/packets per time unit) are characterized by an autocorrelation function that decays *hyperbolically* in the lag (i.e., the lag k autocorrelation behaves like $k^{-\beta}$, for $0 < \beta < 1$). In stark contrast, short-range dependent processes are characterized by *exponentially* decaying autocorrelations (i.e., the lag k autocorrelation decays like ρ^k, for $0 < \rho < 1$. For engineers working in the more familiar frequency domain, long-range dependence corresponds to *$1/f$ noise* behavior, while short-range dependent processes possess spectral densities that remain bounded for low frequencies.

- *Hurst Parameter:* Statistically, self-similarity (or its concomitant long-range dependence) can be checked rigorously in a number of different ways. In increasing degree of statistical sophistication, these tests include variance-time analysis, R/S-analysis using the rescaled adjusted range statistics, and periodogram-based methods such as Whittle's approximate MLE method. As a byproduct, these tests yield an estimate of the *Hurst parameter,* H, that measures the degree of self-similarity or long-range dependence in a given empirical time series. An H-estimate of 0.5 identifies the underlying process that supposedly generated the time series at hand as short-range dependent, while values of the Hurst parameter (strictly) between 0.5 and 1.0 characterize the process as long-range dependent. When applying these tests to the Ethernet LAN traffic data, the outcomes are consistently and convincingly in favor of long-range dependence, irrespective of when and where in the network the traffic data were collected, with typical H-values between 0.75 and 0.95. Interestingly enough, the only data sets that yield H-estimates close to 0.5, indicating that traditional traffic model might be adequate, correspond to actual Ethernet traffic during the not very interesting low activity periods (e.g., night hours), consisting mostly of machine-generated traffic patterns. The Hurst parameter can also be used to describe and measure *burstiness.* This is of practical importance because conventional burstiness measures are typically ill-defined for self-similar traffic, mainly because packet inter-arrival times tend to exhibit extremely high variability which, in turn, can be most efficiently modeled via the *Noah Effect,* i.e., using distributions with infinite variances (see the discussion below).

The implications from this statistical analysis for future traffic modeling research are simple: For gaining a good understanding of how high-speed networks of the future work, learn about the traffic that they are expected to transport in reality.

continued on next page

Self-Similar Traffic Modeling *(continued)*

For learning about the actual nature of this traffic, analyze in detail relevant and representative sets of high-quality traffic measurements (e.g., LAN traffic). and finally, when analyzing large sets of traffic data in a rigorous manner, don't be surprised to discover traffic characteristics (e.g., self-similarity) that defy conventional wisdom.

Why Self-Similar?

One of the first reactions we typically encounter when presenting the findings from our analysis of Ethernet LAN traffic measurements to network and traffic engineers is the desire for a physical or phenomenological explanation for the self-similar nature of aggregate Ethernet LAN traffic. This desire does not necessarily question the results of our statistical analysis but expresses he practitioners' need for an intuitively appealing and mathematically rigorous argument that involves individual Ethernet hosts and relates to practical experience with actual network behavior. The apparent contradiction between the empirically observed self-similarity property of aggregated Ethernet traffic and the supposedly universal applicability of the folklore theorem mentioned earlier in this article is further motivation for identifying the main cause of self-similarity in aggregate packet traffic.

The desired physical explanation is provided by a simple variant of a construction originally due to Mandelbrot [8]. In its simplest form, this variation of Mandelbrot's construction states that if individual Ethernet sources can be described by simple "on-off" models (i.e., during "on" periods, packets are generated in regular intervals, and no packets are generated during the "off" periods), where the lengths of the "on" and "off" periods are drawn (independently) from distributions that have *infinite variance* (i.e., exhibit the *infinite variance syndrome*), then aggregating many such sources produces traffic that is self-similar in the limit (as the number of sources increases). Clearly, this convergence result relies heavily on the infinite variance assumption on the distributions of the lengths of the "on" and "off" periods. Intuitively, infinite variance means "extreme variability" or "variability on all time scales," and captures the property that with non-negligible probability, the "on" and "off" periods can last for very long times. Mandelbrot refers to this property as *The Noah Effect,* in reference to the Biblical story of Noah and the "Big Flood."

In our context, the Noah Effect is what violates the assumptions of Palm's Theorem; hence, it represents the distinguishing feature between aggregate traffic that is self-similar and superposition processes for which the folklore theorem mentioned earlier applies. In fact, the superposition of traditional "on-off" sources (i.e., the distributions of the corresponding "on" and "off" periods have finite variance, or more specifically, are assumed to be exponential) results in a traditional, i.e., short-range dependent traffic process. Moreover, we recently revisited the Ethernet LAN data for a detailed statistical analysis at the source level, and our preliminary results show strong empirical evidence in favor of Mandelbrot's construction and the Noah Effect exhibited by each individual Ethernet source.

A more detailed account of our analysis of the Ethernet data at the source level and a rigorous treatment of the above-mentioned variant of Mandelbrot's construction is currently in preparation [9].

Why self-similarity matters

While we concentrate in this article on Ethernet LAN traffic, self-similarity or long-range dependence appears in measured traffic from a number of different working packet networks—aggregate Ethernet traffic both within local workgroups and across corporate boundaries [5], variable-bit-rate (VBR) video traffic over ATM (for a wide variety of codecs and scenes) [10,11], remote terminal users on ISDN packet networks [12], and signaling networks (CCS/SS7) for the public telephone network [13]. Given that long-range dependence is ubiquitous in measured traffic from today's packet networks and cannot be captured by the currently employed theoretical models for packet traffic, there has been increasing concern about the practical relevance of traditional traffic models and the validity of traffic engineering guidelines that are based on these models. However, only direct arguments detailing the impact of long-range dependence on performance will convince teletraffic and queueing specialists of the value of self-similar traffic models for network design and performance analysis. Thus, while network practitioners typically have no problems in accepting the fact that actual network traffic behaves very differently from what the theoretical traffic models predict, but would like to know why traffic looks self-similar, queueing and performance analysts are much more reluctant. Before taking self-similar traffic for real, they generally demand examples that show why self-similarity matters for queueing.

Clearly, the challenge with the latter response is that queueing theory with long-range dependent input processes represents a new area of research and is likely to require a new set of mathematical tools. Nevertheless, first results in this area are starting to appear and show unmistakenly that the performance of queuing models with long-range dependent input processes can be drastically different from the performance obtained using traditional short-range dependent models. For example, buffer sizing based on traditional models when the actual traffic exhibits long-range dependence, typically results in overly optimistic QOS guarantees. This observation follows from [14] and fully explains the poor performance experienced in first-generation ATM switches (see our earlier discussions and [1]). Also, the results in [14] provide a formal argument for an observation that is well-known to network practitioners, namely that for connectionless traffic, the link utilization cannot be practically improved by simply providing more and more buffers. In fact, while a traditional network engineering rule-of-thumb says that "reducing the free capacity by half can be accomplished by doubling the buffer size," it can be shown that in the presence of long-range dependence, halving the free capacity is only possible by increasing the amount of buffer space by orders of magnitude! These and other results only seem to scratch the surface, and it is safe to expect more dramatic and surprising departures from traditional queueing results as the research efforts into understanding queueing models with input processes that exhibit the Joseph and the Noah Effects increase.

In this context, there is another recent development worth noticing. It attempts to merge the analytical or numerical tractability of traditional queueing models with the practical relevance and statistical accuracy of long-range dependent input streams. The basic idea consists of using traditional traffic models (which are amendable to numerical computations) with a large number of parameters to "imitate" long-range dependence, or equivalently, to approximate the spectral density at the low end of the spectrum.

continued on next page

Self-Similar Traffic Modeling *(continued)*

Following this approach, recent results obtained by Li [15] provide convincing evidence that long-range dependence (or equivalently, the low frequencies) can completely dominate queueing behavior. Clearly, these findings invalidate many currently used traffic engineering rules that are based on traditional models that cannot account for the presence of the Joseph Effect (or the low frequencies).

Keep it simple

Collectively, the results obtained using our proposed "data-driven" traffic modeling approach and the recent findings on how the Joseph and Noah Effects impact performance of high-speed packet networks leads to the next question: how to actually model network traffic accurately and realistically? In other words, what are traffic models that capture the same characteristics observed in measured network traffic? Do these empirically observed characteristics mean "the end of simple traffic models" (see [16])? To answer these questions, recall that by a *traffic characteristic,* we mean a property that has been uncovered by a rigorous statistical analysis of measured traffic data and (i) can be estimated, (ii) can be given a meaningful physical interpretation in the context of the network where the data were collected, and (iii) has been demonstrated (analytically or through simulations) to have a dominant impact on problems related to network engineering. This definition strongly argues in favor of models that obey the *Principle of Parsimony,* also known as *Occam's Razor* [17]. A modern-day rendering of Occam's original principal can be expressed as "An explanation of the facts should be no more complicated than necessary" or "Simple is beautiful."

Somewhat surprisingly, traditional traffic modeling has practically abandoned parsimony as a requirement for traffic models, with the result that it has become commonly accepted in the traffic modeling literature to consider processes with 10 and more parameters. Unfortunately, characterizing traffic using such models yields little insight into the true nature of the traffic, stands no chance of satisfying the three requirements mentioned above, and is of limited practical use (with the possible exception for obtaining numerical results). It turns out that there exist long-range dependent processes where one parameter (namely the Hurst parameter mentioned earlier) suffices to capture the Joseph Effect observed in measured traffic data. These processes are known as *Fractional Gaussian Noise* (FGN) and *Fractional ARIMA Models* and form the basis for self-similar traffic modeling for tomorrow's high-speed networks. For example, LAN traffic can be successfully modeled using a FGN process with three parameters: mean, variance, and Hurst parameter; fractional ARIMA processes with 4 or 5 parameters seem to describe VBR video traffic reasonably accurately. Thus, observing traffic characteristics such as self-similarity, the Joseph Effect or the Noah Effect in a given set of traffic data implies neither "the end of simple traffic models" nor "the beginning of complicated traffic models." The above-mentioned "fractal" traffic models *are* simple, yet they are very different from what we are used to.

Conclusions

Traditionally, traffic modelers use actual traffic traces (at best) to fit the parameters of some predetermined mathematical model. Our analysis of Ethernet LAN traffic measurements (and traffic traces from other working packet networks) demonstrates that when the data are also used for determining the basic characteristics of the traffic, a traffic modeler's life becomes more exciting and surprises are abound.

As very recent studies of WAN traffic (at Lawrence Berkeley Laboratory [18]), NSFNET traffic (Georgia Institute of Technology [19]) and non-Ethernet LAN traffic (Stanford) illustrate, traffic characteristics such as self-similarity, the Joseph Effect, and the Noah Effect are here to stay. In fact, they can be safely expected to play an increasingly important role in the traffic engineering work for tomorrow's high-speed networks.

Given the ubiquitous nature of these novel traffic characteristics, the *good news* is that there exist mathematical models that are able to capture these features accurately and in a parsimonious manner. Moreover, parameter estimation techniques for many of these models are available and algorithms for synthetic traffic generation from these models are known. The *bad news* is that both the estimation methods and traffic generation algorithms are often not yet fully understood and, at this stage, still require careful and often extensive testing and experimenting. This situation impedes, for example, full-scaled simulation studies at this time, mainly because the quality of the traffic generation algorithms is, in general, unknown. Part of the bad news is also that the research work on long-range dependent queueing models has only just started. It will take time and require a coordinated research effort to advance queueing theory for long-range dependent (or self-similar) arrival streams in order to build a "National Information Superhighway" of similar quality to what we expect from today's telephone network (whose QOS and reliability is in no small part due to a mature and established queueing theory for Poisson arrival processes). However, during this process, the continued collection and analysis of actual traffic measurements as the networks and the services they carry evolve over time has to remain an absolute prerequisite. As some statisticians like to say: "Data worth collecting are also worth analyzing"; we might add: "Data worth collecting are worth being used for building future high-speed networks."

References

[1] *Communications Week,* May 16, 1994.

[2] A. Ya. Khinchin, *Mathematical Methods in the Theory of Queueing,* Griffin, London, 1960.

[3] M. F. Neuts, "A versatile Markovian point process," *The Annals of Applied Probability,* No. 16, pp. 764–779, 1979.

[4] D. M. Lucantoni, "New results on the single server queue with a batch Markovian arrival process," *Stochastic Processes and their Applications,* No. 7, pp. 1–46, 1991.

[5] P. F. Pawlita, "Two decades of data traffic measurements: A survey of published results, experiences and applicability," Proceedings of ITC-12, Torino, Italy, 1988.

[6] W. E. Leland, M. S. Taqqu, W. Willinger and D. V. Wilson, "On the self-similar nature of Ethernet traffic," Proceedings of SIGCOMM '93, San Francisco, CA, pp. 183–193, 1993.

[7] W. E. Leland, M. S. Taqqu, W. Willinger and D. V. Wilson, "On the self-similar nature of Ethernet traffic (extended version)," *IEEE/ACM Transactions on Networking,* No. 2, pp. 1–15, 1994.

[8] W. E. Leland, M. S. Taqqu, W. Willinger and D. V. Wilson, "Self-similarity in high-speed packet traffic: Analysis and modeling of Ethernet traffic measurements," *Statistical Science,* 1994 (to appear).

continued on next page

Self-Similar Traffic Modeling *(continued)*

[9] W. Willinger, M.S. Taqqu, R. Sherman and D. V. Wilson, "Analysis of Ethernet traffic measurements at the source level," in preparation, 1994.

[10] J. Beran, R. Sherman, M. S. Taqqu and W. Willinger, "Long-range dependence in variable-bit-rate video traffic," submitted to *IEEE Transactions on Communications,* 1993.

[11] M. W. Garrett and W. Willinger, "Analysis, modeling and generation of self-similar VBR video traffic," Proceedings of SIGCOMM '94, London, UK, pp. 269–280, 1994.

[12] A. Erramilli and W. Willinger, "Fractal properties in packet traffic measurements," Proceedings of Regional ITC Seminar, St. Petersburg, Russia, pp. 144–158, 1993.

[13] D. E. Duffy, A. A. McIntosh, M. Rosenstein and W. Willinger, "Statistical analysis of CCSN/SS7 traffic data from working CCS subnetworks," *IEEE Journal of Selected Areas in Communication,* No. 12, pp. 544–551, 1994.

[14] I. Norros, "A storage model with self-similar input," *Queueing Systems,* 1994 (to appear).

[15] S.-Q. Li and C. L. Hwang, "Queue response to input correlation functions: Continuous spectral analysis," *IEEE/ACM Transactions on Networking,* No. 1, pp. 678–692, 1993.

[16] C. Partridge, "The end of simple traffic models," (Editor's Note), *IEEE Network,* September 1993.

[17] W. H. Jefferys and J. O. Berger, "Ockham's Razor and Baysian analysis," *American Scientist,* No. 80, pp. 64–72, 1992.

[18] V. Paxson and S. Floyd, "Wide-area traffic: The failure of Poisson modeling," Proceedings of SIGCOMM '94, London, UK, pp. 257–268, 1994.

[19] S. Klivansky, A. Mukherjee and C. Song, "Factors contributing to self-similarity over NSFNET," preprint, 1994.

WALTER WILLINGER received the Diplom (Dipl. Math.) in 1980 from the ETH Zürich, Switzerland, and the M.S. and Ph.D. from Cornell University. He is a Member of Technical Staff at Bellcore, where he works in the Computing and Communications Research Department. He is currently an Associate Editor for *The Annals of Applied Probability.* His e-mail address is: `walter@bellcore.com`

DANIEL V. WILSON received his M.S. in Electrical Engineering from Stanford University in 1983 and his B.S. in Physics and Mathematics from Southwest Missouri State University in 1977. He is a Member of Technical Staff at Bellcore where he works on network monitoring and analysis. His e-mail address is: `dvw@bellcore.com`

MURAD S. TAQQU received his B.A. in Mathematics and Physics in 1966 from the Universite de Lausanne-Ecole Polytechnique and his Ph.D. in Statistics in 1972 from Columbia University, New York. Since 1985, he has been Professor in the Department of Mathematics at Boston University. He is currently an Associate Editor for *Stochastic Processes and their Applications* and coauthor of the book *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance* (Chapman and Hall, 1994). His e-mail address is: `murad@math.bu.edu`

WILL E. LELAND received his Ph.D in computer science from the University of Wisconsin, Madison. He is a Member of Technical Staff at Bellcore, where he works in the Network Systems Research Department. E-mail: `wel@bellcore.com`

[W. Willinger and M.S. Taqqu were partially supported by the NSF grant NCR-9404931.]

Networked Voice Messaging

by Greg Vaudreuil, Octel Communications

Overview

A special class of messaging protocols has evolved over the past ten years to meet the needs for *voice messaging*. Voice messaging provides messaging services similar to that of other interpersonal messaging such as electronic mail. Historically these protocols have focused almost exclusively on the requirements for the efficient and economical transmission of high quality voice over the switched telephone network. As digital networking facilities proliferate within corporations, and as electronic messaging evolves to handle mixed media, a convergence of these technologies will occur.

This article provides a brief overview of analog voice networking protocols and provides a description of the *Multi-Media Internet Mail Extensions* (MIME) profile for voice messaging.

History of voice messaging

Voice messaging originated in the telephony industry as an application for voice processing machines, special purpose computers designed to provide telephone answering and *Interactive Voice Response* (IVR) services. These machines typically have a high speed backplane connecting a series of telephony interface units to internal disk drives. The origins in the early eighties of these machines generally required proprietary hardware and operating systems to handle the high-volume real-time data requirements as well as the pre-DSP analog telephony interfaces.

The telephone interfaces have the capabilities to digitize voice in an efficient manner, and to generate and recognize *Dual Tone Multi-Frequency* (DTMF) or *Touch Tone*™ inband signaling. DTMF was implemented to enable a caller to enter data after a telephone connection is established.

The VMX Corporation (Now a part of Octel Communications) first developed voice messaging in 1980. Voice messaging was simply the ability of one user to send a message to another subscriber without first calling their telephone extension. This application was first deployed as a single system solution in 1980. It was expanded to enable the sending of messages to subscribers on other machines, through the development of analog voice messaging in 1985.

Analog messaging is provided by many voice mail vendors. Almost all vendors provide networking between their platforms using proprietary DTMF signaling protocols. In 1987 a group of users and vendors formed an industry consortium to develop the *Audio Messaging Interexchange Specification* (AMIS). This group developed two specifications, one a simple analog DTMF based protocol, and the other a digital protocol based on 1984 X.400.

The final version of AMIS was published in 1991. The analog protocol has seen widespread adoption in the industry and is nearly ubiquitous. The digital version has seen limited implementation.

Analog networking technology

Analog networking has had many advantages of digital networking including simplicity and the ubiquitous availability of analog telephone lines. The drawbacks of analog networking was primarily the degradation of the message from repeated transmissions and repeated analog-to-digital conversions. As the telephony network becomes increasingly all-digital and the analog-to-digital algorithms improve, these limitations are becoming less significant.

continued on next page

Networked Voice Messaging *(continued)*

Minus the overhead of signaling the sender and recipient via DTMF, analog networking provides a 1-to-1 line usage, that is, it takes one minute to transmit a one minute message. This message is sent with the maximum fidelity possible on an analog telephone line. Today with the best high-speed modems and the best analog-to-digital compression algorithms, 1-to-1 line usage is possible, and that is with degraded voice quality. Without fully-digital end-to-end facilities, analog networking still offers the best voice messaging service.

Analog networking uses DTMF "packets" to communicate the sender, recipient, and any selected messaging features. These packets provide integrity checking via a checksum. The following is an example message frame from the industry standard AMIS protocol. Each unit is a single DTMF symbol.

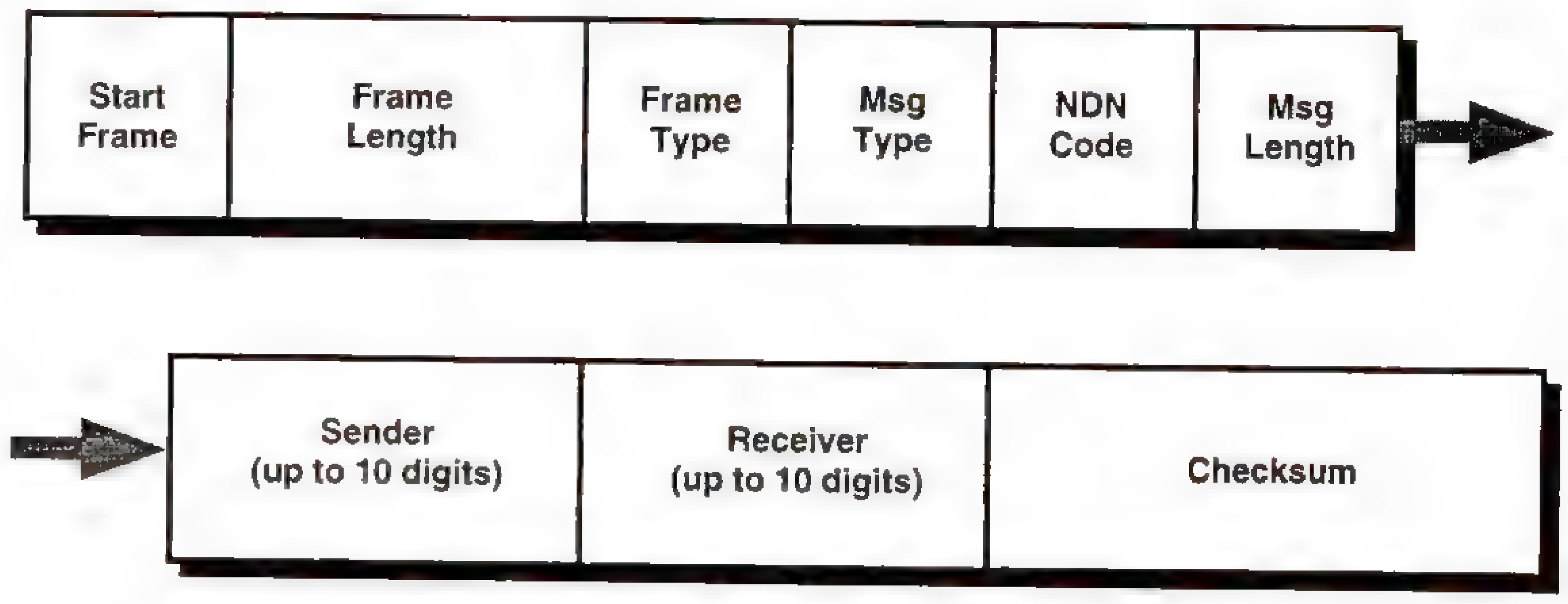

Figure 1: AMIS Message Frame

The frame begins with an asterisk, followed by a base-10 frame length, from 3 to 99 DTMF symbols, and the frame type (new message, start session, end session). If it is a new message frame, the header is followed by the message type (new, reply, or error), the non-delivery notification reason (0 if it is a new or reply message) and an approximate length of the message (up to 9 minutes). The frame is terminated by a checksum optimized for detecting DTMF recognition errors.

The message frame is followed by a response frame containing a response code. If the response is positive, the voice message is played out as an analog signal, terminated by a DTMF pound sign (#).

Other proprietary protocols include per-message and per-recipient delivery options including urgent, private, sending timestamp, and read receipt confirmation.

MIME based voice messaging

Digital networking is increasingly demanded by large voice mail users to reduce costs and increase the quality of service. When digital facilities are available, message transmission time can be significantly reduced while preserving fidelity even with standard audio compression algorithms. A one minute voice message can be sent in 30 seconds over a 64Kbps link using standard 32Kbps ADPCM encodings. Over a T-1 link, the transmission times may be reduced to the point where inter-machine messaging is not significantly slower than messaging intra-machine.

AMIS-Digital was written to use X.400 transport to realize digital transmission. The protocols are based on the 1984 standard and use a bilaterally defined bodypart for the voice message content. Addressing is defined to use 10 digit numbers as a *Domain Defined Attribute* (DDA). Routing is not discussed and is a continuing problem. CCITT has defined a new standard, X.440, for voice messaging based on the 1988 X.400 standard. To date there has been minimal implementation of this approach.

There is increasing interest in using native TCP/IP transport and MIME/ESMTP messaging protocols, in particular for the wide area routing services provided. Corporate networks generally have a deployed TCP/IP network with DNS message routing services. These services can be leveraged to build a large-scale messaging infrastructure. MIME/ESMTP is generally available on a wide variety of multi-media workstations and its use in voice message transport opens the possibilities of distributed voice mail integrated with existing messaging.

To meet the demand for a MIME/SMTP based protocol as well as standard integration other messaging technology issues, a follow-on group to AMIS, the *Universal Message Interoperability Group* (UMIG) was formed. This group is in the exploratory phase of re-incorporating within the *Electronic Messaging Association* (EMA) to gain a broader user base and contribute a fresh perspective to the universal mailbox discussions.

The MIME Voice Profile

The effort to use MIME for voice messaging has resulted in a profile document for MIME and ESMTP. MIME provides a general packaging mechanism for mixed media, but does not define or constrain the range content-types permissible. RFC 822 provides a rich addressing syntax, but user addresses need to be limited to that which can be entered with 10 digit keypads.

There are several missing areas of technology essential for the use of Internet mail protocols for voice messaging. This technology is under development in the IETF and is required in the current voice profile. Current error reports are unusable for automated processing on behalf of a user without a text display device. Machine parsable error reports and extensions to the ESMTP message transport layer to assure their reliable use need to be defined. Binary transport is essential for economical carriage of large binary objects. While Base-64 is a workable solution for casual mixed-media messages, the efficiency loss from that encoding is unacceptable.

Read receipts are a normal part of the voice messaging culture and will need to be defined. The Internet community has historically been reluctant to standardize their use, but a standard reporting format must be defined to enable those who which to use them the ability to interwork.

Next steps

The voice profile document is currently an Internet Draft. The profile depends upon several protocol extensions which are or soon will be on the standards track. After these extension documents are published as proposed standards, the voice profile will be submitted as a proposed standard.

References

[1] Vaudreuil, G., "MIME: Multi-Media, Multi-Lingual Extensions for RFC 822 Based Electronic Mail," *ConneXions,* Volume 6, No. 9, September 1992.

[2] Crocker, D., "Standard for the format of ARPA Internet Text Messages," RFC 822, August 1982.

[3] The MIME RFCs: RFC 1341–1345, June 1992.

GREG VAUDREUIL is a system architect for Octel Communication Corporation, a supplier of telephone answering and voice messaging equipment. His work includes developing MIME based digital voice networking and interoperability standards for the voice messaging community. He served as both the chairman of RFC 822 extensions working group which developed the MIME specification, and the SMTP extensions working group. He has a BS in both Electrical Engineering and Public Policy from Duke University. E-mail: `Greg.Vaudreuil@alfred.ons.octel.com`

Call for Papers

The 6th *IFIP International Conference on High Performance Networking,* HPN '95, will be held in Palma de Mallorca, Balearic Islands, Spain, September 11–15, 1995. The conference aims at presenting and discussing evolution of communications in the framework of high-speed networking and computing in private and public networks.

Topics

Original contributions on the following topics are solicited:

- *New MAC Services and Protocols:*
  Gigabit networks
  ATM-based systems and networks
  LAN emulation on ATM networks
  QoS routing for ATM networks
- *Enhanced Network and Transport Services and Protocols:*
  Multipeer services and protocols
  Admission and congestion control
  Time-constraint management
- *New Services and Protocols:*
  Synchronization semantic and management
  Protocols for groupware communication
  Video over high speed networks
  QoS semantic
- *New applications:*
  Architectural frameworks for distributed multimedia
  Testbeds, implementation experiments and experiences
  Distribution network algorithms
  Groupware communication
  Video conferencing
- *Internetworking:*
  Routing in high performance multimedia networks
  Bridges and routers technology and protocols
  Meshed architectures
- *Implementation and Performance Evaluation:*
  Performance of high speed networks
  Efficient Protocol Implementation
  New implementation architectures

Submissions

Papers must be written in English and should not exceed 12 single spaced pages, or 20 double spaced pages. The front page should contain the authors names, address, phones, faxes, and e-mail, as well as a 150 words abstract. All submitted papers that scope with the topics will be refereed. Authors must join a letter with their commitment to present the paper at the conference. Authors of accepted papers will be requested to sign a copyright release from IFIP. Five copies of the submitted papers should be sent to:

Ramon Puigjaner
Universitat de les Illes Balears
Departament de Ciencies Matematiques i Informatica
Carretera de Valldemossa km. 7.6
07071 Palma, SPAIN

Phone:	+34-71-173288 (direct)	+34-71-173401 (secretary)
Fax:	+34-71-173003	E-mail: `putxi@ps.uib.es`

Important dates

Tutorial Submission Deadline:	November 30, 1994
Paper Submissions Deadline:	January 31, 1995
Notification of Acceptance:	April 30, 1995
Camera-ready copy due:	June 15, 1995

Announcement and Preliminary Call for Papers

INET '95, the 5th Annual Conference of the Internet Society focusing on worldwide issues of Internet networking will be held 28–30 June 1995 in Honolulu, Hawaii. The goal of this conference is to provide a platform that will bring together those developing and implementing Internet networks, technologies, applications, and policies worldwide for infrastructure development. A full call for papers will be published shortly.

Topics

Possible topics for paper submissions include but are not limited to the following:

- Network Technology
- Network Engineering and Operation
- Application Technology
- Users
- Regional Issues
- Policy Issues
- Commercial and Business Aspects
- Education

Submissions

The official language of the conference is English. Papers will be selected based on two-page extended abstracts. The abstracts should include name, address, affiliation, phone number, fax number, and e-mail address. They should also include a keyword list, tied to the track topics listed above. Upon acceptance, full papers will be required. Instructions for preparation of these papers will be provided upon acceptance. Extended abstracts in plain ASCII text should be submitted *by 13 January 1995* to: `inet-submission@interop.com`

The Program Committee can be contacted at:

`inet-program@interop.com`

Workshops

The INET '95 Conference will be preceded by a seven day program of intensive instruction with a hands-on emphasis on Internet set-up, operations, maintenance and management. For information and general questions about the workshop, please send e-mail to: `workshop-info@isoc.org`.

INET '95 will also be preceded by a tentative two day program bringing together active K–12 (Kindergarten through Secondary School) Internet innovators from around the world to share experiences and learn new advanced tools and collaboration techniques. For information and general questions about the K–12 workshop, send e-mail to: `inet-k12-request@isoc.org`

More information

Information concerning the conference is available from the Internet Society Secretariat:

URLs:	`http://www.isoc.org/inet95.html`
	`gopher://gopher.isoc.org/11/isoc/inet95`
	`ftp://ftp.isoc.org/isoc/inet95`

E-mail:	`inet95@isoc.org`
Phone:	+1 703 648 9888
Fax:	+1 703 648 9887
Address:	Internet Society Secretariat 12020 Sunrise Valley Drive Suite 270 Reston, VA 22091 USA

TERENA: Trans-European Research and Education Networking Association

A new initiative to promote a Global Information Society

Amsterdam 20 October 1994—Europe has recently launched the concept of the Global Information Society. According to the Bangeman Report: "Preparing Europeans for the advent of the information society is a priority task. Education, training and promotion will necessarily play a central role."

The concept of the "Electronic Superhighway," has spread all over the globe. Increasingly the European public at large has been informed, and is becoming more and more interested. Individuals want to become actively involved in this new world called the Internet.

RARE + EARN = TERENA

Building on this trend two European organizations—RARE (*Réseaux Associés pour la Recherche Européenne*) and EARN (*European Academic and Research Network*)—have been working towards this ideal. For ten years these two organisations have been actively promoting networking standardization, organizing technical working groups, providing network services, coordinating international projects and promoting the interests of all European service providers for the Academic and Research community. Now, RARE and EARN have decided to merge into one new organization: *TERENA*.

Goals

TERENA's aim is to "promote and participate in the development of a high quality international information and telecommunications infrastructure for the benefit of research and education." To achieve this goal, TERENA will:

- Work towards the elimination of technical barriers by the coordination of standards and operational procedures and the free exchange of technical information
- Provide education, documentation and support services for users of international networks
- Coordinate improvements of international communications traffic
- Organize conferences, meetings, and workshops to promote and improve international networking
- Provide round table discussions with governments, standard bodies, telecommunications operators and industry
- Undertake projects to develop new pan-European services required by the membership

Already, RARE and EARN individually have strong track records in these areas. TERENA has been established to expand these activities further and to widen participation in European networking. Together, RARE and EARN can develop a high-speed networking infrastructure that will bring Europe effortlessly into the information world.

Membership

TERENA's membership at its inception on 20 October consists of representative organizations from 38 countries and two International Treaty organizations (CERN and ECMWF).

The combined strengths of RARE and EARN will be augmented by well-known international corporations, for instance IBM and Digital Equipment have agreed to join TERENA as Associate Members to help make TERENA the powerhouse for development of European research networking.

Softbank Corporation Buys ZD Expos

Foster City, CA (November 2, 1994)—Softbank Exposition and Conference Company, formerly Ziff Davis Exposition and Conference Company (ZD Expos), the trade show company for the 21st century, announced its purchase by Softbank Corporation of Tokyo, Japan for $202 million. The company, will continue to be based in Foster City and run by its current management team headed by president, William Lohse. The 1995 calendar of trade show events, featuring domestic and international occurrences of *NetWorld+Interop, Seybold Seminars, Digital World* and *Windows Solutions,* will remain as scheduled.

Content based tradeshows

The exposition company quickly established itself as a major force in computer-related trade shows and conferences. In 1994, its revenues are projected to exceed $90 million, more than four times its 1991 revenues. The success of these events begins with a high quality conference that attracts the best and brightest in the community. The interactive trade show floor is a professionally enriching experience for the qualified attendees who visit the exhibits of the industry's key players. Each event is extended by a year-round, industry focused newsletter and, most recently, the development of an on-line virtual trade show.

"Softbank is an ideal fit for us," said Lohse. "They recognize and support the key components of our success in creating 21st century trade show events."

"Well established shows and its tremendous growth potential drove our decision to buy," said Masayoshi Son, president and chief executive officer of Softbank. "They have proven themselves a forward looking company that is constantly looking to provide increasingly better service to their customers."

Events

NetWorld+Interop is the world's leading networking event. From the desktop to the enterprise to the Internet, NetWorld+Interop delivers the information that LAN, WAN and Telecom buyers need to make global decisions. More than 200,000 people attended NetWorld+Interop events worldwide in 1994 in Las Vegas, Atlanta, Tokyo, Berlin and Paris. The company recently introduced *N+I Online!*, the virtual trade show. *N+I Online!* provides online access to the wealth of information and events, before, during and after NetWorld+Interop events.

Seybold Seminars, held in San Francisco, Boston and Paris, are recognized as the world's premier computer publishing and graphics events. Windows Solutions, held in San Francisco, Frankfurt, Tokyo and Paris, explores the boundaries of client/server Windows applications for developers and system integrators worldwide. Digital World is the premier gathering place for digital media professionals and is devoted to products that advance the state of the art.

Newsletters

In order to extend the educational influence of its trade show events, Softbank Exposition and Conference Company publishes six reports that are recognized as "sources of record" in their specialties: *ConneXions—The Interoperability Report* for NetWorld+Interop; *Windows Watcher* for Windows Solutions; *Digital Media* for Digital World; and *The Seybold Report on Desktop Publishing* and *The Seybold Report on Publishing Systems* for Seybold Seminars.

CONNEXIONS
303 Vintage Park Drive
Suite 201
Foster City, CA 94404-1138
Phone: 415-578-6900
FAX: 415-525-0194

FIRST CLASS MAIL
U.S. POSTAGE
PAID
SAN JOSE, CA
PERMIT NO. 1

ADDRESS CORRECTION
REQUESTED

CONNEXIONS